La recherche d'informations sur internet

Licence CC BY

Comment bien rechercher une information sur internet.

#ZdS

Auteurs : elyppire933 Eskimon pierre_24

Devant la multiplicité des sources d'informations et la soif d'apprendre de tous les zesteurs, voici un petit guide de "comment trouver les réponses à mes questions avant d'aller demander sur un forum". Nous traiterons ainsi de divers sujets partant de la base des choses en parlant des moteurs de recherche pour finir vers un brin de pro-activité avec la veille.

Bien que cela puisse paraître évident à certaines personnes, la recherche d'informations possède son lot d'astuces et de choses à connaître pour être optimale. C'est un peu comme la cueillette des champignons, il faut connaître les bons coins .

Vous pourrez tirer de ce tutoriel plusieurs avantages. D'une part une nouvelle tranche de culture et d'autre part une compétence supplémentaire, celle de ne plus avoir besoin d'attendre des réponses à vos questions sur un forum puisque vous pourrez trouver par vous-même en un clin d'œil !

- Les moteurs de recherche
- Les autres sources d'informations
- La recherche dans le cadre des sciences
- La veille technologique

Les moteurs de recherche

Comment ça marche ?

Un *moteur de recherche* est une application vous permettant de retrouver des éléments (pages web, vidéos, images, ...) associés aux termes de votre requête. Ils ne sont pas limités au domaine de l'internet et peuvent également prendre la forme de programmes installables directement sur votre ordinateur, vous permettant alors d'effectuer des recherches parmi les documents présents sur votre ordinateur.

Cependant, nous allons nous contenter ici de parler de la recherche d'informations sur le web dans le cadre des sciences et de l'informatique.

Le fonctionnement d'un moteur de recherche se compose en général de trois étapes :

- Une phase dite *d'exploration*, effectuée en général par un robot, qu'on nomme aussi *crawler*. Le rôle de celui-ci est d'explorer de manière systématique le web en suivant en général les liens présents sur les ressources explorées.
- Une phase dite *d'indexation* : de ce qui a été récupéré par le *crawler*, il s'agit alors d'extraire les termes pertinents afin de les ranger dans une importante base de données. Différents algorithmes sont employés dans cette étape afin de donner un poids aux mots les plus importants du document.
- Une phase de *recherche* : il s'agit de la phase où vous intervenez ! C'est la phase où le moteur fait la traduction de votre requête en résultats issus de sa base de données, qu'il vous restitue. Une fois encore, de nombreux algorithmes peuvent intervenir dans cette phase, afin que le moteur classe ces résultats selon l'ordre qu'il trouve le plus pertinent (c'est le cas du célèbre PageRank de Google).

À noter que la plupart des moteurs de recherche actuels, en plus de donner du sens au(x) mot(s), cherchent désormais à comprendre la structure du document qu'ils analysent, on parle dès lors de *sémantique*. Comparez ça à l'effet que possède un mot en gras dans une phrase, si on écrit "le **chat** est dehors" et "le chat est **dehors**", bien qu'il s'agisse de la même phrase, vous sentez bien que l'information qui est mise en avant n'est pas la même, ce qui modifie le sens de la phrase. C'est ce que désigne le terme de sémantique, et c'est pourquoi il faut utiliser du formatage de texte avec parcimonie : écrire un texte entièrement en gras sera peut-être intéressant visuellement, mais ne met rien du tout en avant, et un moteur de recherche ne fera alors pas la différence entre les termes.

La recherche "simple"

Prenons un exemple très terre-à-terre et qui pourrait très bien arriver dans la vie réelle. Il existe, en PHP, plusieurs fonctions permettant d'afficher des informations à l'écran : echo, print(), print_r(), var_dump(), ... Vous venez de lire un tutoriel parlant de ces fonctions et vous trouvez que la différence entre elles n'est pas très claire. Plutôt que de poster un message sur le forum PHP, vous avez la bonne idée de faire une recherche auparavant. Comment ?

Bien entendu, vous pourriez ouvrir votre moteur de recherche préféré et y entrer simplement print() ou echo. Le problème est que vous risquez de vous retrouver avec des résultats qui n'ont rien à voir avec ce que vous cherchez : ces fonctions existent dans plusieurs langages, et les mots *print* et *echo* possèdent eux-mêmes une signification. Vous risquez donc de devoir lire plusieurs pages de résultats inutiles avant de trouver l'information qui vous convient. Et si vous êtes pressés, ça ne risque pas d'être efficace (n'hésitez pas à essayer si vous ne me croyez pas). Il

est donc important de savoir ce que l'on cherche et surtout de **choisir les termes de sa recherche** … Sans tomber dans l'autre extrême ! Ainsi taper "quelle est la différence entre echo() et print() en php ?" risque également de vous donner le bon résultat, pour peu que quelqu'un se soit posé la question avant vous. En effet, le moteur de recherche ne comprend généralement pas le sens des mots et se bornera dès lors à les rechercher dans sa base de données (certains moteurs de recherche sont toutefois équipés pour reconnaitre et comprendre de manière plus ou moins efficace le langage humain).

En conclusion , **une bonne requête doit être la plus courte et la plus ciblée possible**. La meilleure requête étant celle qui, avec le minimum de caractères, vous permettrait d'obtenir le meilleur résultat correspondant à votre attente. Dans ce cas différence echo print php devrait amplement suffire (les puristes feront d'ailleurs remarquer que cette différence n'existe pas dans d'autres langages et que le terme "php" est dès lors inutile).

À noter que les moteurs de recherche ne tiennent généralement pas compte des majuscules ou des accents, taper "RÊVE" ou "reve" renverra donc généralement les mêmes résultats. Le singulier ou le pluriel les changeront par contre, puisque le moteur de recherche pense en général qu'il s'agit d'un mot différent, de même pour le masculin et le féminin. C'est particulièrement flagrant sur des mots dont le pluriel n'est pas que le simple ajout du suffixe "s" au singulier : essayez sur des mots tels que "journal" ou "bocal".

Améliorer votre recherche

Comme nous l'avons dit, un moteur de recherche n'est généralement pas équipé pour comprendre le langage humain, d'ailleurs, avouons-le, assez complexe. Cependant, il reconnait tout de même certains mots ou signes qui ont, pour lui, une importance particulière.

1. Le premier est le signe "+", qu'on retrouve également sous la forme de "AND" (qui signifie "ET" en français). Il est en quelque sens implicite, puisque le moteur de recherche les rajoutera en général pour vous, et signifie que vous recherchez plusieurs termes en même temps. Ainsi, la requête donnée plus haut, différence echo print pourrait être réécrite différence+echo+print. Mais la différence entre les deux ne sera en général pas visible, puisque comme il a été dit, le moteur de recherche les rajoute automatiquement.
2. Le second est le signe "-", qu'on retrouve aussi sous la forme de "NOT" (signifiant "SANS" en français). Comme on pourrait s'y attendre, il permet d'exclure des résultats. Par exemple, imaginez que vous voulez exclure les résultats issus de Zeste de Savoir, parce que vous y avez déjà fait une recherche et que vous n'avez pas trouvé votre réponse, il suffit d'écrire -"zestedesavoir" (remarquez l'usage des guillemets dans ce cas, qui sera rediscuté plus loin). Cela permettra de filtrer les résultats contenant le terme "zestedesavoir" (il existe par ailleurs un moyen plus "propre" d'obtenir le même résultat, mais qui sera expliqué plus loin). Cette fonction revêt beaucoup plus d'importance dans d'autres situations : par exemple, faire une recherche sur le mot "trigger" dans le contexte du javascript renvoie à sa signification dans le contexte de la célèbre bibliothèque jQuery, dont la popularité fait que les moteurs de recherche supposent généralement qu'il s'agit du but de votre recherche. Enlever ces résultats avec -jquery devient alors un véritable atout.
3. Le troisième terme est "OR" (signifiant "OU" en français). Assez logiquement, il permet de rechercher un terme OU un autre. Reprenons l'exemple des fonctions en PHP : vous venez d'apprendre que print_r() et var_dump() sont des *alias* l'un de l'autre (pour ceux à qui ces termes ne diraient absolument rien, il s'agit tout simplement qu'écrire l'un ou l'autre ne change rien au résultat, ils sont synonymes). En conséquent, si vous voulez connaître la différence entre ces fonctions et echo, une bonne requête serait différence echo (print_r OR var_dump) (remarquez l'usage des parenthèses, qui sera expliqué plus loin), qui donnerait des résultats différents (plus complets) que la simple requête différence echo print_r.
4. Le joker : "*". Il s'agit tout simplement de *remplir les vides*. Mettons que vous ne vous intéressez pas à la différence entre print et print_r, mais que vous voulez garantir un maximum de résultats, vous écrirez print*. Ainsi, le moteur de recherche prendra en compte les résultats comprenant les deux termes. La différence est beaucoup plus visible sur des recherches comparatives entre animal et anima*, qui reprendra également les termes "animaux", "animalerie" et ainsi de suite, mais aussi "animateur", "animation", …

Tous ces termes sont nommés *opérateurs booléens*, en référence à un grand nom de la logique, George Boole, qui a donné son nom à cette logique. D'autres opérateurs existent, aux fonctions différentes : XOR, NAND, … Mais ne sont pas forcément supportés par les moteurs de recherche, et sont généralement moins employés dans le domaine de la recherche d'information (on laisse ça aux programmeurs). Par ailleurs, il existe des moteurs de recherche où le "-" n'a pas d'effet et où il faut employer "NOT". La francisation de ces termes peut également être plutôt mauvaise, et reste à essayer au cas par cas.

Les parenthèses permettent de mieux scinder les expressions de manière logique et en modifie les résultats. Un exemple ? Écrire "(jambon OU fromage) ET beurre" signifierait que vous avez le choix entre le jambon et le fromage et que vous avez d'office le beurre. Au contraire "jambon OU (fromage ET beurre)" signifie que vous avez le choix entre le jambon (mais sans beurre) et le fromage (mais avec du beurre). Les parenthèses sont donc très importantes, même si dans la pratique, on les emploie généralement peu (encore un truc de programmeurs).

Enfin, dans les exemples qui précédent, il a été fait usage dans certains cas de guillemets. Ils permettent de dire au moteur de recherche de non pas chercher chaque terme séparément, mais de manière groupée. Et ainsi, rechercher "différence echo print" n'aura pas du tout le même effet que la requête plus générale, puisque le moteur de recherche fera dès lors en sorte de mettre en avant les résultats où le terme est employé tel quel (ou, à défaut, les plus proches possibles). Remarque intéressante, on devrait alors tomber sur ce tutoriel, qui ne répond absolument pas à la question, preuve que bien choisir sa requête est important. On n'emploie donc les guillemets que pour une recherche sur un terme exact ou une expression bien précise (on force alors le moteur de recherche à trouver là une chaîne de caractère).

La recherche avancée

Il existe également certains opérateurs spécifiques, qui peuvent être ajoutés en fin de recherche afin de restreindre les résultats de manière plus efficace qu'à l'aide d'une combinaison absconde de mot clés et d'opérateurs. Il en existe un grand nombre, qui changent en fonction du moteur de recherche employé. Il sont de la forme <mot-clé>:<valeur>. Une liste très réduite est présentée ici :

1. Recherche sur un site en particulier : il suffit de préciser site:<adresse du site> à la fin de la requête. Ainsi, pour faire une recherche sur Zeste de Savoir, il suffira d'écrire site:zestedesavoir.com. Bien entendu, il est possible de préciser plusieurs fois cet opérateur afin de faire la recherche sur plusieurs sites. Et mieux encore, il est également possible de le faire précéder d'un "-", afin d'au contraire supprimer les résultats issus d'un site (exemple : -site:wikipedia.org).
2. Rechercher un type de document précis : ext:<extension> ou filetype:<type>. C'est parfois utile afin de rechercher uniquement un type précis de document (PDF, doc, Excel, …), ou des images d'un format particulier.
3. Rechercher les pages qui pointent vers un lien : link:<adresse du lien>. Permet d'obtenir les pages pointant sur un lien en particulier. Il peut s'agir d'un site en général ou d'une page en particulier.

Comme expliqué auparavant, il est loin de s'agir des seuls. Mais comme vous savez désormais faire des recherches performantes et précises, ne tient qu'à vous de les trouver et de les employer le cas échéant

Des exemples de moteurs de recherche

La liste des moteurs de recherche sur internet est loin de pouvoir se compter sur les doigts d'une main (ou alors, il faut plusieurs mains et beaucoup de doigts). En faire une liste exhaustive relèverait donc du suicide et n'aurait que peu d'intérêt. En voici donc certains parmi les plus utilisés/connus :

Un bon moteur pour une bonne recherche

- Google : sûrement le plus connu et le plus employé. Chéri et décrié, adoré ou détesté, chacun possède son idée sur la question. Il reste cependant incontournable, moteur de recherche par défaut dans un grand nombre de navigateurs internet, et propose en outre plusieurs sous-moteurs de recherche spécifiques : images, actualité, ...
- Bing : concurrent direct du précédent, lancé par Microsoft et *de facto* moteur de recherche par défaut d'Internet Explorer, anciennement *Live Search*. Il offre également plusieurs sous-moteurs de recherche pour les images, vidéos et autres. Il s'agit dès lors d'une alternative viable au précédent, bien qu'il soit plus récent.
- Yahoo! : le moteur de recherche numéro un du temps où Google n'était encore qu'un projet dans la tête de ses créateurs, et qui se défend comme il peut face à ses concurrents. Il s'agissait à la base d'un annuaire de site web, qui possède encore aujourd'hui certains *fans* irréductibles. Il possède également un grand nombre de services complémentaires pour l'actualité, les images et ainsi de suite.
- DuckDuckGo : derrière ce nom quelque peut étrange se cache un moteur de recherche plutôt étonnant, puisqu'il part du principe qu'une requête doit donner le même résultat quelque soit l'utilisateur, au contraire d'autres systèmes de recherches. Il possède également une fonction intéressante, les "!bang", qui permettent d'effectuer des recherches directement sur différents sites : "!g" permettra d'effectuer la recherche sur Google par exemple)
- Qwant : Nouveau venu dans le vaste monde des moteurs de recherche, voici Qwant, un moteur français ! Sa particularité repose sur l'affichage des résultats directement en différentes catégorie, quelques soit la requête. On retrouve par exemple les résultats Web (classique) mais encore "Social", "Actualités"...
- Wolfram|Alpha : *Outsider* dans le monde de la recherche, Wolfram Alpha propose de traiter des requêtes en langages naturels. Ici, pas d'hyperlien ou quoi que ce soit dans les résultats. Le but est de vous donner une information factuelle. Pour bien faire vos requêtes, il faudra donc formuler une question en anglais pour ensuite avoir diverses réponses (mathématiques, statistiques, historiques...). Essayez, c'est très intéressant !

Le moteur de recherche de Zeste de Savoir

Zeste de Savoir possède un outil pour faire des recherches sur tout le contenu du site ou seulement une partie. Il est basé sur un moteur de recherche open source nommé Solr. La base de toute recherche sur Zeste de Savoir doit donc commencer par ici. En effet, dites-vous bien que la plupart du temps vous n'êtes sûrement pas le premier à affronter le problème auquel vous faites face.

Cette recherche préalable vous évitera de poser des questions maintes fois traitées, qui risqueraient de provoquer des remarques désobligeantes de la part de membres las de se répéter.

Cet outil de recherche personnalisée fourni par la fondation Apache est gratuit et open source. Vous pouvez vous en servir pour créer votre propre moteur de recherche pour votre site web, par exemple. Attention cependant pas de miracles, ce moteur de recherche nécessite tout de même quelque efforts de configuration pour fonctionner correctement.

Les autres sources d'informations

Les sites officiels

Qui est plus au courant d'une technologie que son créateur lui-même ? De nos jours, toute création d'une taille conséquente possède sa documentation : allant de celle pour les bibliothèque dans les différents langages à celle des fonctionnalités de base du langage que vous étudiez, en passant par les manuels de fonctionnement de programmes en ligne de commande. Cette documentation présente la plupart du temps des exemples ainsi que des explications supplémentaires sur certains paramètres : ce domaine profite comme tous de l'avantage d'internet, c'est à dire les liens hypertextes qui permettent de relier de manière plus efficace que n'importe quel livre ou tutoriel les informations complémentaires. N'hésitez donc jamais à consulter la documentation de votre langage ou bibliothèque préférée avant de venir poser une question, la réponse s'y trouve souvent !

Wikipédia

Wikipédia se définit lui-même comme un "projet d'encyclopédie multilingue collaborative universelle et librement diffusable". C'est ce qui en fait sa force et sa faiblesse en même temps.

Ainsi, cette encyclopédie se construit au fur et à mesure des contributions de tout un chacun, car tout le monde peut (et est même fortement invité à) apporter sa pierre à l'édifice : le site est dit en évolution continue ! Sa popularité en fait un des carrefours incontournables de la toile, et grand nombre de moteurs de recherche vous proposeront probablement Wikipédia comme réponse pour un certain nombre de mots clés (on pourrait d'ailleurs croire qu'il s'agit d'un partenariat, il n'en est rien). Ce site couvre un nombre tout bonnement incroyable de sujets, de la pomme à des algorithmes très complexes en passant par des biographies ou des oeuvres d'art. Le projet s'est d'ailleurs diversifié, puisqu'il propose également des cours (la Wikiversité) ou un suivi de l'actualité (Wikinews).

Les critiques qu'on peut adresser à Wikipédia sont nombreuses, et certaines sont justifiées (elles sont reprises par Wikipédia sur une page dédiée, dont la neutralité est laissée à l'appréciation de chacun). On peut par exemple penser à la facilité d'un vandalisme. La communauté autour du site ne laisse en général pas passer ce genre d'action, mais certaines modifications peuvent passer inaperçues durant plusieurs semaines voire mois. Ces erreurs sont heureusement minimes par rapport à la quantité d'informations. On peut également pointer du doigt le manque de connaissances des contributeurs dans certains domaines : en effet, puisque le contenu est amélioré par les utilisateurs, ce sont les connaissances des utilisateurs qui sont utilisées. Il n'y a pas de confirmation de la part d'autorités reconnues, puisque tout se fait dans l'anonymat relatif à internet.

Le logo de Wikipédia

Encore d'autres sources !

Les méta-moteurs de recherche

Il s'agit d'un moteur de recherche qui exploite les résultats d'autres moteurs de recherche. L'avantage est qu'il est généralement capable d'en extraire des différences et d'éliminer les résultats similaires. De plus, les utiliser permet de tronquer l'effet de mise en avant de certains sites par des moteurs de recherche. Enfin, ils proposent parfois des options avancées de tri des résultats. Il peuvent être employés directement dans un navigateur, ou sont parfois téléchargeables à part et utilisables en GUI. Il sont cependant très peu utilisés.

Les annuaires de liens

Au contraire d'un moteur de recherche, l'Indexation de liens est réalisée par des êtres humains. En conséquent, les résultats obtenus sont, en comparaison, bien moindres, mais d'une qualité généralement supérieure. Cependant, ces annuaires, très employés aux débuts de l'Internet ont perdus de leur intérêt face à la facilité d'utilisation des moteurs de recherche. Ils sont généralement employés aujourd'hui comme aide au référencement. Yahoo! était à la base un annuaire de lien, aujourd'hui reconverti en moteur de recherche.

La recherche dans le cadre des sciences

Les sources d'informations en sciences

On pense généralement à internet quand il s'agit de faire une recherche, et c'est d'ailleurs le but de ce tutoriel, mais ce n'est pas la seule source. Demandez à vos parents, lorsqu'il s'agissait de faire un exposé, c'était direction la bibliothèque. Aujourd'hui, un grand nombre de ces documents a été numérisé et est disponible, pour des coûts variables, sur le net. De manière générale, on trouve, sur le net, différents médias apportant de l'information :

- Les pages web ;
- Les documents informatiques (PDF, présentations, ...) ;
- Le livre numérisé (légalement ou non) ;
- Les publications issues de journaux scientifiques, qui peuvent prendre un certain nombre de formes en fonction de leur origine et leur but : *proceeding*, *review*, ... ;
- Les thèses de doctorat ou rapport de master de recherche.

Chacun de ces types de sources à ses qualités et ses défauts, ainsi que ses spécificités : une publication ne dépassera généralement pas la vingtaine de page, alors que c'est un minimum attendu pour un livre. Quant à une page web, bien qu'elle puisse atteindre de manière théorique une longueur infinie, sera en général assez courte. Par ailleurs, on attend généralement d'un livre qu'il parcoure un sujet de manière plus approfondie et plus large, ce qui signifie qu'il faut du temps pour écrire un livre : au contraire d'autres médias, un livre traitera de sujets de fond, alors qu'une publication se rapporte, au moment de sa publication, tout du moins, à une recherche qui est d'actualité (sauf s'il s'agit d'un *review*). Une thèse de doctorat aura généralement la même vocation.

Il faut garder en tête cette notion de temporalité quand on recherche une information. Et bien entendu, un livre ou un article aura été relu par plusieurs personnes au moment de sa publication, il est donc généralement moins susceptible de contenir des informations incorrectes. Mais la temporalité peut prendre plusieurs formes : n'oubliez pas qu'on a longtemps cru que la terre était plate. Autrement dit, plus la source est vieille, plus il y a de chances qu'elle ait été remise en question.

Internet : ne croyez pas tout ce que vous voyez !

Internet, c'est un cadeau et un poison . Car c'est un fait avéré sur internet dans tout les domaines : on peut y raconter n'importe quoi, du très vrai comme du très faux. À l'heure où une personne avec un minimum de connaissances peut créer un site web, à l'heure de la démocratisation des réseaux sociaux, l'information va vite, très vite. Et toutes les sources semblent se valoir. Et pourtant, une information partagée par un blog à 100 lecteurs par mois aura-t-elle autant de valeur que si elle était annoncée par un quotidien reconnu ? *Ça dépend*. Il est donc nécessaire de se poser quelques questions.

Bien sur, vérifier la véracité d'une information, n'est pas nécessaire tout les jours. Il existe même des cas ou nous pouvons nous en passer. Mais, et plus particulièrement en sciences, une valeur ou un résultat peuvent changer beaucoup de choses, ne serait-ce que la réponse à un exercice. Quelques précautions sont donc parfois nécessaires. Mais la règle générale reste *méfiez vous de tout*.

- **L'adresse du site** : elle peut vous donner beaucoup d'informations, à commencer par l'extension : en effet, les sites en ".gov" sont obligatoirement tenus par des gouvernements, et les sites en ".edu" par des organismes éducatifs reconnus, tandis que n'importe qui peut posséder une extension en ".com". De la même manière, l'adresse du site d'une université contiendra en général les initiales "ac" ou le terme "univ". Au contraire des extensions de domaines, qui sont gérées par des organismes sérieux, il ne s'agit là que d'une "spécification", qui n'est pas toujours respectée. De manière générale, pensez toujours à remonter à la racine du site (l'adresse commençant par "http" et s'arrêtant au premier "/") afin de voir où est hébergé l'information que vous êtes en train de consulter. Repérez également les adresses contenant des "~" : elles sont souvent le signe d'un espace personnel sur un serveur général, bien que cette règle tend également à être de moins en moins respectée avec la démocratisation des sous-domaines.
- **Le site** : L'habit ne fait pas le moine, dit le proverbe, mais sur internet, tendez à penser le contraire. En effet, une information aura tendance à être plus crédible sur un site crédible que sur un site qui paraît fait à la va-vite. C'est pareil avec un livre : on préférera un livre qui reste sobre qu'un livre qui change de couleurs toutes les 5 lignes pour trouver une information. Puis ça fait moins mal à la tête. Attention ! Ce n'est pas parce que le site semble sérieux que l'information l'est également. Certains sites sont sciemment créés dans ce but (l'exemple le plus connu reste le site du DHMO, je vous laisse trouver pourquoi).
- **L'auteur** : qui a écrit ce que vous êtes en train de lire ? Forcément, un être humain qui a voulu que vous le lisiez. Vérifiez que cet auteur donne son nom. Et s'il ne le fait pas, pourquoi ? Est-ce parce qu'il s'agit d'un site commercial, et que l'auteur parle au nom de sa société, ou par oubli ? Si l'auteur est mentionné, est-il connu ? Est-il compétent dans le domaine dans lequel il donne l'information que vous cherchez. N'hésitez jamais à faire des recherches complémentaires sur des noms d'auteurs ou des sociétés, elles permettent parfois de discriminer des résultats par manque de sérieux.
- **L'information elle-même** : vous semble-t-elle cohérente ? En général, à moins que vous ne fassiez des recherches sur un domaine de pointe, plusieurs personnes pensent généralement la même chose. Si une information semble en désaccord avec d'autres, il semblerait une bonne idée de se demander pourquoi. Est-ce que c'est ce que l'auteur cherche à montrer, ou ce qu'il tient pour acquis. Explique-t-il le pourquoi de cette différence avec d'autres auteurs ?
- **Les sources** : le principe de base en sciences est d'avancer en s'appuyant sur les recherches de ses collègues. De ce fait, la citation est une activité très importante : certains jugent même la qualité d'une publication au nombre de ces sources. Sans tomber dans de tels extrêmes, regardez si l'information que vous avez sous les yeux est corroborée par des citations. Elles vous permettront de gagner un temps précieux en permettant de rapidement contrôler la validité d'une information en allant consulter ce que d'autres auteurs en pensent (mais attention à l'auto-citation, pratique courante dans le milieu) et vous indiqueront si l'auteur s'est documenté sur le sujet ou pas. Bien entendu, ces sources peuvent être de toutes sortes : livres, articles, autre page internet, … Et bien entendu, chacune d'entre elle doit également faire preuve d'une critique de votre part.
- Les remarques données plus hauts concernant la temporalité de l'information sont évidement toujours à prendre en compte.

Au delà de toutes ces règles plus empiriques qu'autres choses, l'important est de réfléchir à ce que vous êtes en train de lire. Si cela vous semble mauvais, vérifiez ! Il en va de la crédibilité de ce que vous dites, car il n'y a rien de plus décrédibilisant que de citer de fausses informations, surtout si quelqu'un s'en rend compte.

À propos de Wikipédia

Puisque, comme expliqué plus haut, tout un chacun est en mesure de modifier une page de cette encyclopédie sans que cela soit forcément validé par la communauté, généralement par manque de connaissances, on préfère dès lors, **éviter de citer Wikipédia** en général. Ce qui ne veut pas dire qu'il ne faut pas l'employer ! Ce projet reste tout de même un moyen facile d'obtenir l'accès à une information, dont il faudra ensuite recouper la véracité avec d'autres sources. En ce sens, Wikipédia permet d'agir comme annuaire de citations, vu qu'il propose souvent des sources considérées comme "acceptables" sur le sujet présenté sur la page.

Une croyance populaire veut qu'un scientifique doit bouder Wikipédia. Que nenni ! Il est le plus à même de l'améliorer, en particulier dans des domaines de pointes où Wikipédia manque encore de précision.

Les moteurs de recherche spécifiques

Là où Google (ou un autre) ne vous donnera que des résultats issus de l'ensemble du web (bien qu'il soit possible d'employer des filtres, voir plus haut), certains moteurs de recherche sont dits "spécialisés" et indexent uniquement certains types de contenu. Ces moteurs existent pour un grand nombre de domaines, en fonction de la volonté de leur(s) développeur(s) et de l'intérêt que les gens y portent. Le domaine scientifique ne fait pas exception : les contenus indexés par ce genre de moteurs sont de types académiques ou universitaires. On y retrouve des ressources issues de sites en *.edu, des bases de données d'éditeurs de revues, des catalogues de certaines bibliothèques, des ressources libres d'accès, des brevets ou encore des bases de données bibliographiques. Le tout est filtré de manière plus ou moins complète. Bien entendu, il est toujours possible d'affiner la recherche avec des mots-clés bien choisis et des opérateurs logiques.

À noter que dans le domaine scientifique, une recherche aura souvent plus de résultats si elle est effectuée en anglais. Et de manière générale, cette règle s'applique également pour les opérateurs booléens : *OR*, *AND*, *NOT*, … Mais dans le cadre des sciences, la connaissance de l'anglais reste un plus incontestable, même pour effectuer des recherches d'informations.

Moteurs multi-disciplinaires

Le premier d'entre tous n'est pas trop dépaysant, puisqu'il s'agit d'une initiative *made in Google* : il s'agit de **Google Scholar** : vous pouvez donc profiter de l'expertise du géant Google dans le domaine de la recherche informatique bien que les résultats soient mieux filtrés qu'avec une (longue) série d'opérations logiques. Google Scholar possède un grand nombre d'avantages, mais on peut lui reprocher de ne pas proposer d'options avancées quand à la recherche sur un auteur ou sur une source précise (un mode "avancé" existe heureusement pour pallier à ces quelques désagréments). De plus, l'algorithme de classement n'est pas connu, et nous n'avons aucune information sur la période couverte. Ce moteur de recherche propose en outre quelques fonctionnalités avancées intéressantes, telles que la possibilité de créer des alertes par e-mail.

Le second est nommé **Mendeley** : là où Google proposait encore une interface en français, il s'agit cette fois d'un site entièrement anglophone. Il est produit par Elsevier, société détenant un certain nombre de journaux scientifiques assez connus dans leurs domaines. De ce fait, ce moteur couvre un nombre étendu de domaines, de la physique théorique pure aux sciences sociales et économiques. Son domaine d'indexation est également un peu plus complet que la liste présentée plus haut, puisqu'il s'intéresse également aux pages gouvernementales (.gov) sur les domaines scientifiques et à certaines pages de conférences scientifiques. Son interface de recherche en mode avancé est également beaucoup plus complète et propose un nombre impressionnant d'options et fonctionnalités qui raviront certains. Il est aussi possible d'exporter des références des résultats obtenus sous différentes formes (dont le célèbre BibTeX) afin de les archiver ou de les utiliser. Mendeley est également un programme de gestion de documents très pratique, et entièrement gratuit.

Si l'anglais vous rebute et si payer pour avoir accès à des documents n'est pas dans vos cordes, vous pouvez également essayer **isidore**, un moteur de recherche français qui se concentre sur les données disponibles en accès libre mais, malgré tout, à visées scientifiques : pages universitaires (mais aussi écoles secondaires), documents numérisés issus de bibliothèques scientifiques, … Le domaine couvert est une fois encore assez large, bien que le nombre de documents y soit malgré tout limité et que le moteur se concentre plus spécialement sur les sciences humaines et sociales. Il s'agit d'un projet développé entre autre par le FNRS, qui doit encore connaître plusieurs évolutions.

Il en existe encore d'autres, visant différents buts et plus ou moins orientés : http://worldwidescience.org/, http://www.springeropen.com/ (mise à disposition d'articles par une autre société d'édition de journaux scientifiques, Springer), http://www.science.gov/, http://www.openaire.eu/ (issus de la communauté européenne), http://www.freefullpdf.com/ (permet l'accès à un grand nombre d'articles, cependant parfois limité à *l'abstract*), Library Genesis pour télécharger certains contenus plus difficiles d'accès, ou encore http://www.citebase.org/ (qui ne donne accès qu'a des listes de citations, à vous de cherchez le document, et bien sûr, on ne cite pas quelque chose qu'on n'a pas lu).

Moteurs de recherche spécifiques

- Physique : on peut mentionner arXiv, qui,se présentant plutôt comme un site austère, couvre un grand nombre de domaines de la physique (d'ailleurs pas uniquement, on y trouve aussi des publications dans le domaine des mathématiques et de l'informatique). ArXiv (à prononcer comme "archive" à l'anglaise) est un endroit ou un certain nombre de scientifiques prépublient leurs travaux, ce qui donne à ce site une réputation contestée dans le domaine, mais il reste cependant une source d'information intéressante.
- Mathématiques et informatique : ThechXtra est un bon début. Le plus populaire, à ce jour, reste CiteSeer. On retrouve également certains articles sur arXiv.

- Biologie et médecine : BioSearch propose, une fois encore, d'effectuer des recherches dans des ressources libres d'accès, à l'aide d'une interface très complète (possibilité de rechercher des figures). Mais le domaine reste monopolisé par PubMed pour la recherche médicale, ainsi que l'énorme base de données sur le sujet est BioMedCentral, dont deux de ses projets frères ont déja été mentionnés plus haut.
- Chimie : ChemBioFinder, de la société qui produit le très célèbre ChemDraw. Une inscription est demandée, mais la recherche est ensuite gratuite. Il y a aussi Scopus, qui est un moteur de recherche produit par Elsevier, une fois encore. Il propose des options plus avancées que Scirus mais nécessite souvent un accès qu'on obtient seulement dans les université. Même chose pour son rival, SciFinder, produit par L'ACS (*American Chemical Society*), une fois encore, un grand éditeur de journaux scientifiques, qui en fait le moteur de recherche le plus complet dans le domaine de la chimie (à cause de la période de temps traitée qui est juste immense). Nous déplorons toutefois que seul un accès à travers une université ou une société, qui en a payé les droits, soit nécessaire.

N'hésitez pas à en utiliser plusieurs. En effet, ils ne couvrent généralement pas les mêmes catalogues, on obtient donc souvent des résultats sensiblement différents.

La veille technologique

Afin de garder un oeil sur ce qui se passe autour de vous, différents outils existent pour agréger les nouvelles. Ainsi, en seulement quelques minutes (prévoir un peu plus si vous êtes un gourmand de l'information) vous allez pouvoir vous tenir au courant de ce qui se passe autour de vous !

Les sites officiels

En général, obtenir une info sur l'évolution d'une technologie se fait en allant directement à la source de l'information. Bien des développeurs possèdent un blog ou un site dédié sur lesquels ils annoncent l'état d'avancement de leur(s) projet(s), c'est d'ailleurs assez annoncé sur les sites de projet(s) de taille(s conséquentes, ou des membres sont tout spécialement désignés pour faire ce travail. Ils sont souvent employés en parallèle d'un compte tweeter (voir plus bas), l'un renvoyant généralement vers l'autre.

Si le projet sur lequel vous cherchez des informations est un logiciel libre et open source, il y a de fortes chances qu'il soit hébergé sur un site de partage comme SourceForge ou GitHub. Ces sites proposent alors de mettre en lien les différents développeurs, de montrer un wiki de l'application ainsi qu'un bug tracker.

Les réseaux sociaux

Vous en connaissez sûrement plusieurs, les réseaux sociaux sont omniprésents sur internet. Ils sont un moyen efficace de partage de l'information parfois plus rapides que les médias habituels. Qu'ils soient sous la forme de micro-messages sur Twitter ou d'articles complets sur Facebook, ils peuvent être un très bon moyen de se tenir au courant de ce qui se passe autour de vous et de rentrer en contact directement avec une marque, société ou association qui vous intéresse.

Facebook

Logo de Facebook

Facebook propose un système de page pour que des tiers puissent communiquer avec leur communauté. Ils peuvent ainsi donner des nouvelles sur les technologies qu'ils développent ou encore organiser des rencontres ou des discussions. C'est aussi un très bon vecteur pour "prendre la température" par rapport à un évènement.

Facebook peut s'utiliser de différentes manières. Soit vous êtes intéressé pour participer et échanger, auquel cas il vous faudra créer un compte et rejoindre la masse du milliard d'utilisateurs. Soit vous voulez simplement vous tenir au courant et aller lire des informations et dans ce cas là, vous pouvez utiliser le site en "lecture seule".

Dans le premier cas, vous pouvez vous servir de la barre de recherche située en haut du site et visible à tout moment pour rechercher ce qui vous intéresse. Ensuite, en cliquant sur "j'aime", vous pouvez devenir fan d'une page et ainsi recevoir des informations lorsque cette dernière en publie. Si vous n'êtes pas connecté, vous pouvez toujours rechercher des personnes ou des pages. Cependant, facebook est malin et préfère vous voir connecté ! Pour éviter cela, sur la page d'accueil/d'inscription il suffit d'aller tout en bas de la page, dans le footer. Vous trouverez alors deux liens, un pour "Personnes" (si leurs règles de confidentialité le permet) et l'autre pour "Pages". Vous aurez ensuite accès à une boîte de recherche, en haut et à droite, au dessus de l'alphabet. Une fois que vous avez trouvé une page qui vous intéresse, vous pouvez enregistrer l'adresse dans vos favoris pour venir la consulter de temps en temps. Par exemple, essayez de trouver Zeste de Savoir !

Facebook ne contrôle pas les créations de pages. Il se peut donc que certaines soient créées par des fans plutôt que l'entité officielle (fréquent dans le cas des célébrités). Dans ce cas, la véracité des infos doit être prise avec des pincettes.

Twitter

Logo de Twitter

Twitter est communément appelé "site de micro-blogging". En effet, au lieu de publier des articles entiers, les auteurs de "tweets", les micro-messages de 140 caractères que les gens postent, vont exprimer des humeurs, partager des trouvailles intéressantes sous forme de lien hypertexte ou encore *teaser* sur des nouveautés ou faire des demandes d'aides raccourci.

La force du réseau vient des interactions simples mais efficaces mises en place par le site. Un utilisateur peut s'exprimer ("Tweeter") mais aussi faire des retweets en reprenant un message de quelqu'un d'autre pour le publier à son tour et ainsi exprimer son accord avec l'idée émise. L'information peut ainsi se propager très rapidement.

Twitter propose un système de followings/followers qui permet de suivre des personnes/entreprises et inversement. En ayant des contacts en commun, vous verrez les échanges entre ces personnes.

Un système de hashtags (ou "mots-dièse" en français) vous permettent aussi de rejoindre des conversations ou de les suivre plus facilement. Ils sont repérables grâce au dièse (signe #) qui les précède, un peu comme les @mentions, qui sont précédées d'un arobase. Ces dernières vous permettent de vous adresser à un compte (personne, marque, association…) en particulier et faire un lien depuis votre tweet vers le compte concerné. Un tweet commençant par une @mention n'est, par défaut, visible que de façon réduite (auteur, cible et contacts communs éventuels), ainsi la convention veut qu'on précède cet arobase d'un simple point afin de rendre la conversation publique.

Zeste de Savoir a, par exemple, son compte Twitter : @zestedesavoir, ainsi que d'autres membres. Je vous laisse comme exercice le soin de trouver les auteurs de cet article sur ce réseau

Par défaut, un tweet est considéré comme public. Ainsi, pas besoin d'avoir un compte sur Twitter pour aller suivre les personnes qui vous intéressent (mais c'est quand même plus pratique de les centraliser en s'abonnant).

Les flux RSS

Un flux RSS (pour *Really Simple Syndication*) est une ressource dont le contenu est généralement mis à jour de manière automatique selon les modifications d'un site web. Afin de lire ces flux, il est nécéssaires de les ajouter dans un logiciel nommé *agrégateur*. Une fois que c'est fait, c'est ce logiciel qui s'occupera automatiquement d'aller regarder de temps à autre qu'il n'y a rien de nouveau. Il s'agit donc d'un moyen simple et efficace de suivre l'actualité d'un blog, d'un quotidien ou d'un forum, sans qu'il soit nécéssaire de s'y rendre soi-même.

Les sites proposant ce système d'abonnement sont légions. D'ordre général, on peut les repérer facilement en trouvant cette icone qui signifie qu'un flux est disponible sur la page. On peut souvent aussi trouver le symbole dans la barre de navigation du navigateur indiquant la possibilité de s'abonner. Un clic sur cette dernière devrait vous renvoyer une liste de logiciel(s) disponible(s) pour l'utiliser.

Parmi les agrégateurs les plus connus, on en retrouve de nombreux en ligne. Voici une liste non exhaustive des plus fréquentés :

- Feedly
- The Old Reader
- Digg Reader

Il existe aussi des solutions sous forme de GUI. On en retrouve souvent installés dans les logiciels de messagerie comme Thunderbird ou Outlook.

Les inclassables

Beaucoup de ces sources sont disponibles via différents sites ou technologies : RSS, Facebook, Twitter, LinkedIn, etc. N'oubliez pas d'y jeter un œil, il y a parfois des petits plus intéressants.

StackOverflow

Si vous prenez l'habitude de formuler vos questions en anglais dans Google, vous allez vite vous rendre compte que ce site est LA référence des problèmes de programmation. Ce site est une sorte de gros forum ou les gens viennent poster une question et la communauté répond avec plaisir. Les réponses les plus pertinentes sont mises en avant afin d'avoir une meilleure visibilité. Ce sont les lecteurs et donc la communauté qui votent pour une réponse afin de l'élire comme référence. De nombreux langages informatique sont traités, ce qui fait d'elle LA plate forme de référence si vous parlez anglais (ce qui est indispensable dans le monde de l'informatique et des sciences).

Chaînes YouTube

Certaines personnes aiment les supports vidéos et se débrouillent bien avec. De plus, certaines activités se prêtent plutôt bien à cet exercice, notamment dans le cadre des sciences où un phénomène optique sera plus simple à expliquer par une vidéo que par un bloc de texte. On peut aussi retrouver de plus en plus de tutoriels d'informatique, certains "classique" à base de captures d'écran, d'autres plus originaux reprenant les codes de l'école avec un tableau noir et des craies.

On peut également retrouver de nombreuses conférences sur différents domaines, notamment lors de grands événements mondiaux.

Blogs de passionné(e)s

Dans le paysage internet, une grande mode à l'heure actuelle est celle des blogs, sorte de journal de partage. Nécessitant peu de connaissances techniques pour être mis en place, de plus en plus de personnes franchissent le pas pour en créer un et partager ainsi leurs connaissances, tester du matériel ou discuter de sujets plus philosophiques. L'avantage des blogs est qu'il offre la plupart du temps un bouton de syndication RSS. Cela vous permet ainsi de vous tenir au courant plus facilement des nouveautés. La plupart d'entre eux dispose aussi d'une zone de commentaires en dessous des articles pour laisser les lecteurs s'exprimer et réagir.

Il existe des blogs de toutes sortes, parlant de la couture ou des nouvelles technologies. Certains sont à vocation professionnelle et possèdent le même genre de fonctionnement qu'un journal papier (ligne éditoriale, rédacteur en chef, pigiste…), tandis que d'autres sont plus personnels, avec juste un rédacteur qui veut partager sa passion.

Périodiques traditionnels

On en entend rarement parler et pourtant ils existent, ce sont bien sur les journaux et magazines dans leurs versions papiers. Plus question de passionnés qui veulent partager leurs connaissances, ici on parle de documents censés être professionnels, écrits par des gens se qualifiant de professionnels. Cependant, on retrouve toujours la même variété d'informations que l'on pourrait avoir sur le web. Le fait que le support soit physique entraîne logiquement un coût à payer, mais est souvent gage de qualité en ayant des articles courts, mais aussi des articles ou dossiers de fond sur un thème plus complet ou compliqué.

Conclusion

42

(Et si vous ne comprenez pas pourquoi : cliquez-ici !)

Nous espérons que ce tutoriel vous a été utile, qu'il vous a permis de découvrir de nouvelles choses et qu'à l'avenir, avant de poser une question sur le forum, vous aurez d'abord ce réflexe nommé "recherche"

Table des illustrations

- Merci à Piwit pour le logo du tutoriel
- Mashup logos : Sur le blog d'Olivier Corneloup
- Logos : Depuis leurs sites respectifs

Commentaires

STAFF

Kje — Mercredi 16 juillet 2014 à 15h29

Chouette et bien complet. Apres j'ai lu en diagonal mais j'ai trouvé une petite coquille :

> Le premier d'entre tous n'est pas trop dépaysant, puisqu'il s'agit d'une initiative made in Google : il s'agit de **Google Sch oo lar** [...]
>
> *— Section "Moteurs multi-disciplinaires"*

il faut qu'un seul O

Edité mercredi 16 juillet 2014 à 15h30 par Kje

gustavi — Mercredi 16 juillet 2014 à 16h02

De même, dans la conclusion : s/ils/il et le point final.

dosmpm — Mercredi 16 juillet 2014 à 16h12

Les moteurs de recherche->Améliorer votre recherche->point 4

Ca serait mieux de mettre

```
`*`
```

au lieu de

```
"*"
```

sinon tout l'italique se déplace au début de la phrase.

Viho — Mercredi 16 juillet 2014 à 23h52

C'est complet, c'est bien, c'est beau. Bravo. (-:

STAFF
Eskimon Jeudi 17 juillet 2014 à 08h45

Corrections effectuées, merci à vous !

pierre_24 Jeudi 17 juillet 2014 à 09h36

> > Le premier d'entre tous n'est pas trop dépaysant, puisqu'il s'agit d'une initiative made in Google : il s'agit de **Google Sch oo lar** [...]
> > — *Section "Moteurs multi-disciplinaires"*
>
> il faut qu'un seul O
>
> — *Kje*

Je me trompe tout le temps (même en tapant l'adesse). Y'a un machin en anglais qui justifie la disparition de ce "O" ou c'est juste pour le "plaisir du nom" ?

STAFF
Kje Jeudi 17 juillet 2014 à 09h57

> Je me trompe tout le temps (même en tapant l'adesse). Y'a un machin en anglais qui justifie la disparition de ce "O" ou c'est juste pour le "plaisir du nom" ?

Comment ça ? "Scholar" se traduit généralement par "Savant" en français. Ce qui correspond bien à la cible du produit de Google, non ?

pierre_24 Jeudi 17 juillet 2014 à 09h58

Je pensais que c'était la même racine que "school", *mea culpa*

STAFF
Kje Jeudi 17 juillet 2014 à 10h01

Ha mais c'est la même racine, ça vient du latin je crois...

valent20000 Jeudi 17 juillet 2014 à 12h10

+1

Beau tuto ! (Reste plus qu'à apprendre l'anglais au français moyen pour qu'il puisse faire ses recherches de manière optimale) Faudrait que j'en parle a des amis qui sont prof documentalistes ça devrait les intéresser !

Edité jeudi 17 juillet 2014 à 12h10 par valent20000

Aero Jeudi 17 juillet 2014 à 16h06

dans la présentation de yahoo, il est écrit concucrent

Fulbert Jeudi 17 juillet 2014 à 19h29

Il y a plusieurs occurrences de

> tout les *[...]*

Et je pense que certains termes entre guillemets pourraient être en code inline ou entre des guillemets français.

Sinon c'est bien d'avoir un tuto comme ça. J'espère que les débutants le liront.

Vernom Il y a 20 heures

+3

Au vu du titre, j'me suis dis "ouah, qu'est-ce ça fout là ? Un bout de tuto basique histoire de faire de la prévention pour le forum ?", au final, je suis très étonné.

Super complet, j'ai appris quelques nuances et termes.

Merci & beau boulot !

Vous devez être connecté pour pouvoir poster un message.

Sommaire

1. Les moteurs de recherche
2. Les autres sources d'informations
3. La recherche dans le cadre des sciences
4. La veille technologique

Télécharger

HTML

PDF

EPUB

Archive

Zeste de Savoir • 2014 • Version : v1.0-RC5/746d9d0

CGU À propos L'association Contact